



SHARDING

FRAGMENTATION DE BASE DE DONNÉES

Réalisé par :
Elghali Abir , Triki Asma & Mhiri Mariem

Activer Windows
Accédez aux paramètres pour activer Windows.



PLAN

- 1 Introduction
- 2 Définition et implémentation du Sharding
- 3 Les avantages et les inconvénients
- 4 Exemples d'applications du sharding
- 5 Conclusion





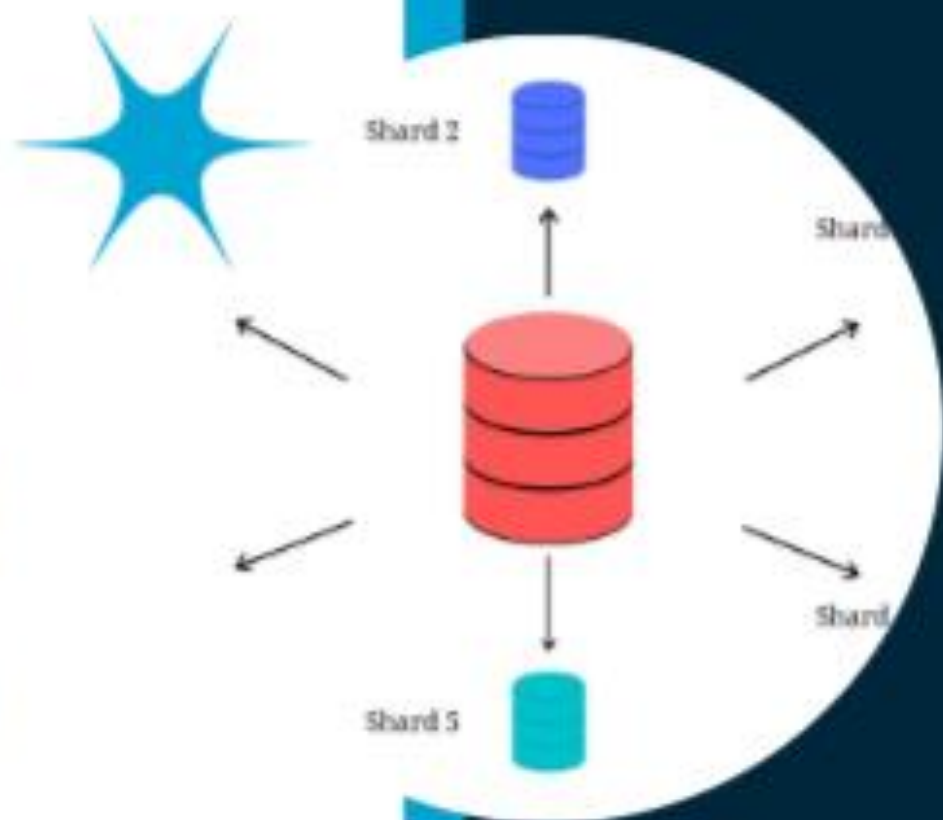
INTRODUCTION

Les systèmes de bases de données comportant de grands ensembles de données et des applications à haut débit peuvent mettre à rude épreuve la capacité d'un serveur unique. Des taux de requêtes élevés peuvent épuiser la capacité



Définition du Sharding

- Le mot « sharding » signifie « éclater ». Dans le domaine de la data, le sharding est une méthode qui permet de partitionner un ensemble de données venant d'une même base de données. On fractionne ainsi notre base de données en plusieurs sous-ensembles de données également appelées « datasets ».





Implémentation du Sharding

- Il existe deux principales méthodes de sharding ou partitionnement : le **partitionnement horizontal** et le **partitionnement vertical**.
- Le partitionnement horizontal permet de **séparer les lignes** et de les stocker dans différents serveurs.
- Le partitionnement vertical quant à lui permet de **séparer les colonnes** et de les stocker dans différents serveurs.



Exemple

Original Table

CUSTOMER ID	FIRST NAME	LAST NAME	FAVORITE COLOR
1	TAEKO	OHNUKI	BLUE
2	O.V.	WRIGHT	GREEN
3	SELDA	BAGCAN	PURPLE
4	JIM	PEPPER	AUBERGINE

Vertical Partitions

VP1

CUSTOMER ID	FIRST NAME	LAST NAME
1	TAEKO	OHNUKI
2	O.V.	WRIGHT
3	SELDA	BAGCAN
4	JIM	PEPPER

VP2

CUSTOMER ID	FAVORITE COLOR
1	BLUE
2	GREEN
3	PURPLE
4	AUBERGINE

Horizontal Partitions

HP1

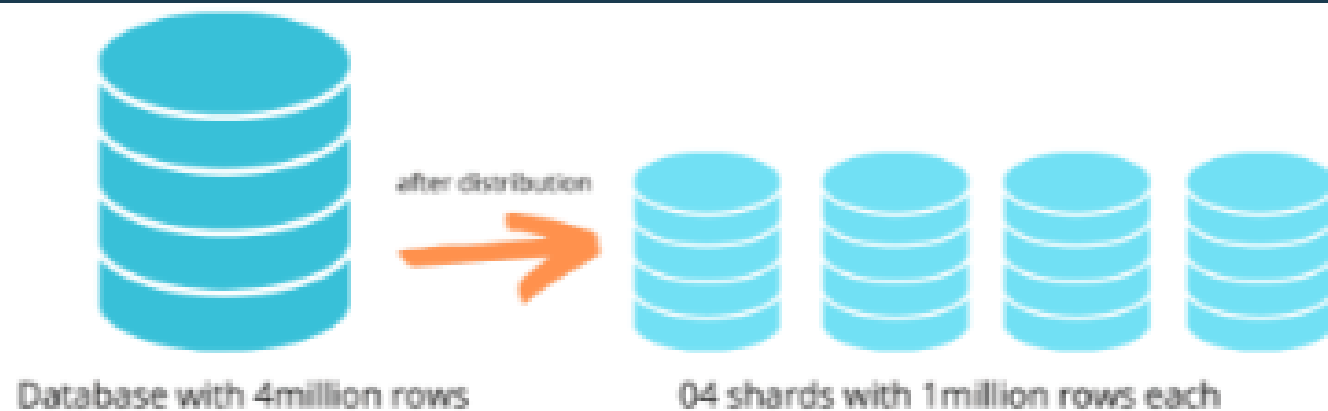
CUSTOMER ID	FIRST NAME	LAST NAME	FAVORITE COLOR
1	TAEKO	OHNUKI	BLUE
2	O.V.	WRIGHT	GREEN

HP2

CUSTOMER ID	FIRST NAME	LAST NAME	FAVORITE COLOR
3	SELDA	BAGCAN	PURPLE
4	JIM	PEPPER	AUBERGINE



Les avantages de Sharding



Le sharding permet de **dépasser la capacité d'une seule machine** en hébergeant les données sur différents serveurs. Ce système de décentralisation possède de nombreux avantages.

- **La sécurité** : c'est plus sécurisant de répartir ses données sur plusieurs serveurs .
- **Des temps de réponse plus rapides aux requêtes**
Les shards ne comportent que quelques lignes et colonnes. De ce fait, le traitement des requêtes de la base de données prend moins de temps.



- **La rapidité** : beaucoup plus d'informations pourront être traitées chaque seconde.
- **Fiabilité accrue dans les situations de panne** : Il minimise les effets des pannes. Comme chaque shard est autonome, seul le shard affecté subit un temps d'arrêt



Les inconvénients de Sharding



- Une mauvaise implémentation peut engendrer une perte de données plus ou moins conséquente qu'on ne pourra plus récupérer

- **Coût supplémentaire** : il peut également entraîner des coûts supplémentaires en matière d'infrastructure, de maintenance et de développement.
- **Gestion des migrations de données** : lorsque la structure de la base de données change, la migration des données peut devenir un processus complexe, surtout si le schéma des données est modifié de manière significative.



Exemples d'applications du sharding



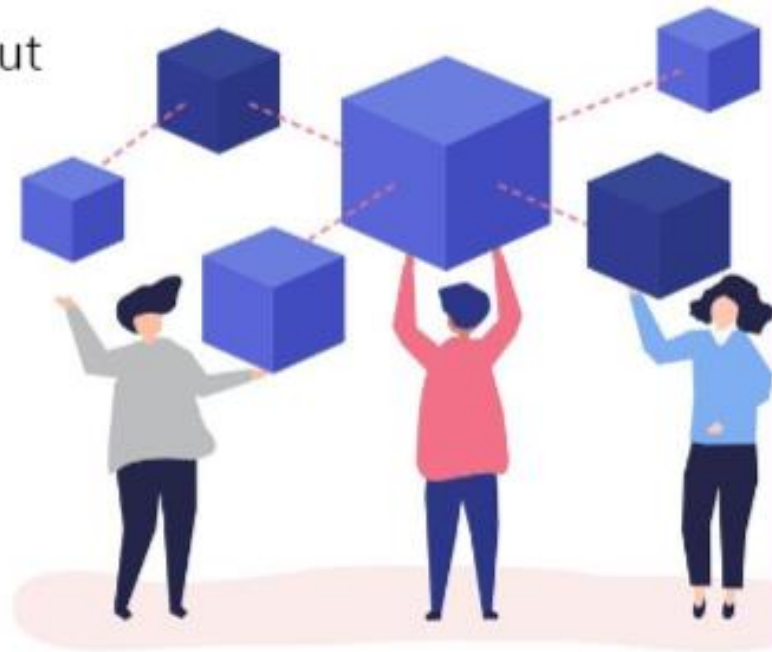
Réseaux Sociaux



Base de données

Blockchain

Dans le contexte des blockchains, le sharding est une technique pour améliorer l'évolutivité en divisant le réseau en sections plus petites appelées "shards". Chaque shard peut traiter ses transactions indépendamment, ce qui accélère le processus global.



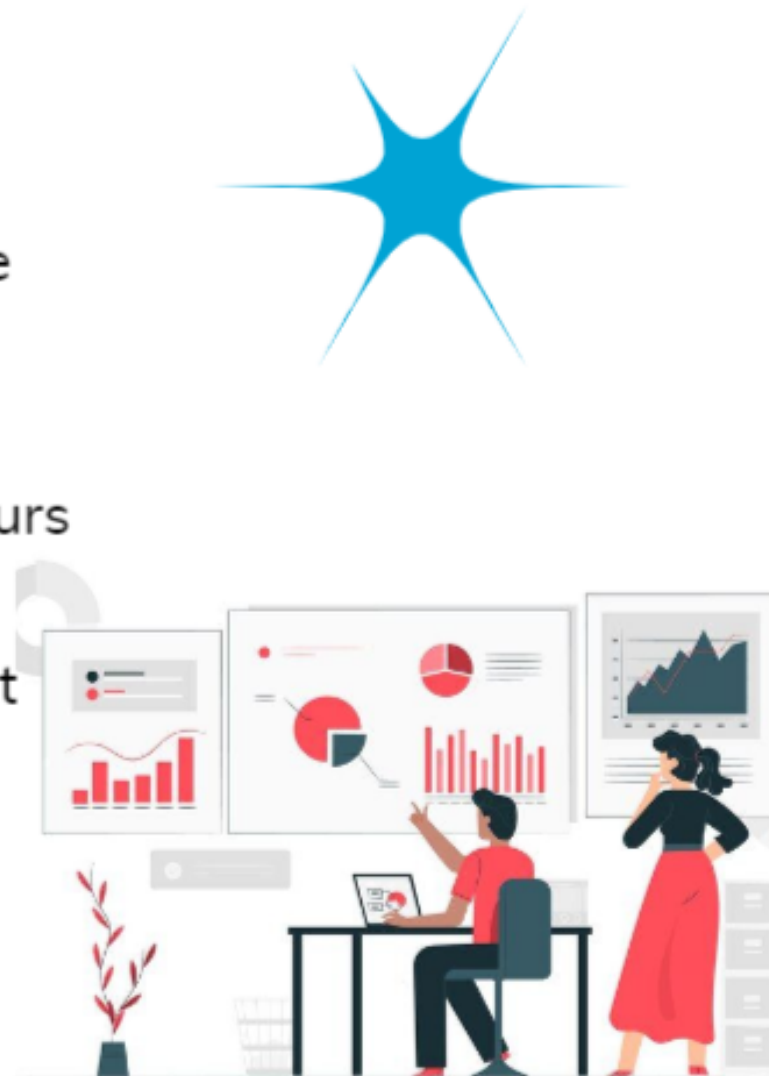
Jeux en ligne massivement multijoueurs (MMOG)

Les jeux en ligne à grande échelle utilisent souvent le sharding pour répartir les joueurs et les données de jeu sur plusieurs serveurs, permettant ainsi une expérience de jeu fluide et réactive.



Analyse de données

Dans le domaine de l'analyse de données et du Big Data, le sharding peut être utilisé pour distribuer les données sur plusieurs nœuds de calcul, facilitant le traitement parallèle et accélérant les opérations analytiques.





CONCLUSION

En résumé, le sharding est une stratégie efficace pour améliorer la scalabilité des bases de données en répartissant les données entre plusieurs serveurs. Bien que cela puisse optimiser les performances, la mise en œuvre doit être soigneusement planifiée pour surmonter les défis liés à la complexité et à la gestion.





Merci pour votre attention



15

Made with VISME